Approaches to the Evaluation of Intelligence

Massive Military Data Fusion and Visualisation: Users Talk with Developers

Dr. Ing. Hans-Joachim Kolb
MEDAV GmbH
Graefenberger Str. 32-34
D-91080 Uttenreuth
GERMANY

hans.kolb@medav.de

Traditionally, intelligence gathering has involved three areas of technology: sensor and signal processing, database technology and evaluation technology. The current pace of globalisation, the rapid penetration of Internet usage and continuing advances in sensor sophistication mean that vast amounts of information are being gathered, as never before. Traditional intelligence processing methods are feeling the strain.

MEDAV GmbH is a developer of intelligence processing hardware and software. For the past 20 years, we have been engaged in studies and commercial projects for the German Federal Armed Forces and for other German and international government agencies. Our business is to predict and respond rapidly to the needs of the intelligence community. The current trends are clear:

1) An increasing variety of sensors are being distributed over larger and larger areas. The problem is not just the amount of information being collected but the diversity of it. The challenge is to find meaningful ways to integrate diverse information. Humans are not as good at data analysis as computers but they are excellent at pattern recognition. The two should be allowed to complement one another. A visualisation tool should present data in a way that allows humans to recognise the patterns within it – as opposed to presenting only pre-defined patterns. Visualisation software that does not allow users to apply their own pattern recognition intelligence is self-defeating.

2) The Internet explosion means that open source intelligence (OSINT) is becoming as important as the traditional HUMINT and SIGINT. Thus we require language processing and document processing software to complement the signal processing software already available. In particular, tools for the visualisation and summarisation of text are required. Another consequence of internationalisation is the necessity to incorporate Unicode capability into software. However the multitude of *ad hoc* solutions to script encoding already in existence will continue to persist in the future despite the adoption of Unicode.

3) To cope with information overload, it will be necessary to integrate more closely the three domains of information acquisition, storage and processing. In particular, automated evaluation will be required closer to the point of acquisition in order to filter at an early stage the mass inflow of irrelevant information. Likewise, storage must be more closely coupled to evaluation technologies in order to give users easy and intuitive opportunities to interact with their data.

4) In the intelligence environment, information is usually incomplete. For example, the resources necessary to resolve ambiguous or missing words may not be available. And context, which allows information to be interpreted, may also be incomplete. Such *cold start* situations are typical in military intelligence. The conclusion is that information processing algorithms, which are able to process incomplete and fuzzy information, will become the norm.

Paper presented at the RTO IST Workshop on "Massive Military Data Fusion and Visualisation: Users Talk with Developers", held in Halden, Norway, 10-13 September 2002, and published in RTO-MP-105.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
00 APR 2004	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Approaches to the Evaluation of Intelligence	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **MEDAV GmbH Graefenberger Str. 32-34 D-91080 Uttenreuth GERMANY**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001665, RTO-MP-105 Massive Military Data Fusion and Visualization: Users Talk with Developers., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**21**	

The rapidly changing needs of the intelligence community place severe constraints on the development of software solutions:

1) Hardware and software must be modular, so that parts can be regularly updated without disrupting the whole. Modules must communicate through standard interfaces. Modularity allows a *building blocks* approach to the construction of complex functionality. Because effective visualisation tools tend to have long development times and therefore high costs, modularisation will be essential.

2) It is better to have generic modules that can be configured by the user for varying situations as they arise, rather than specific solutions for specific situations. The user, rather than the factory, should be able to reconfigure a visualisation tool, for example, so that it operates interactively in one situation but automatically in another. Open architectures and the ability to modify software are preferable to proprietary black-box solutions. However, visualisation tools tend to be task specific. The challenge will be to find ways of developing generic visualisation tools.

3) Solutions to intelligence problems involve compromises. In the real world, it is necessary to make do with what one has. Complex algorithms developed in the ideal world of the research laboratory may not be practical in a world of limited resources. It is easy to be blinded by hi-tech solutions when simpler solutions may be more practical.

4) Many of the problems confronted by the intelligence community are also faced by large commercial operations, for example media and finance organisations. It is better to use COTS tools where these are available, rather than "re-invent the wheel". COTS products make sense if they reduce costs and inventory without compromising quality and security.

Over the past 20 years, MEDAV has developed integrated intelligence processing packages, that acquire, automatically evaluate, archive and interactively evaluate both signals and text documents. The archive is the hub of the architecture. Automatic evaluation serves two purposes - it filters out the mass of irrelevant information but it also serves to enrich the incoming signals or documents with annotations that assist subsequent evaluation.

Different processing and visualisation modules can be incorporated into this general architecture according to need. In studies with the German Federal Armed Forces, we have developed a variety of speech processing, text processing and visualisation tools. While it is true that "a picture tells a thousand words", we find that users need to have different views over the same data. For example a graphical display of a military command structure can be usefully complemented by a simple tabular display of other information. However, neither display by itself is adequate. As another example, a graphical content summary of a text document may be useful if the document is large but for small documents (or parts of documents), the user may prefer textual summaries.

It is helpful to distinguish three types of display; 1) traditional static predefined displays, 2) augmented reality displays (where interactive iconic and textual information are embedded in realistic terrain images, for example) and 3) virtual reality displays (where the user becomes part of the image). We still have much to learn about how to use all three effectively. The development of effective visualisation tools depends on a keen understanding of the psychology of vision. The conclusion is that the results of fundamental research into human visual processing must be fed into software development projects.

SYMPOSIA DISCUSSION – PAPER NO: 14

Author Name: Dr. Ing. Hans-Joachim Kolb, MEDAV, Germany

Question:
What statistical pattern recognition is being used in regards to email?

Author's Response:
No features concerning content of the messages are being used, just behavior, such as time elapsed from getting an email to answering it, or the size of the message.

Question:
Is there any special storage or computing being used for the high volume?

Author's Response:
No. If you calculate how much space you need, you find you can get a great deal of data on one disk. If you are successful in finding an area of interest in an image it makes sense just to keep the area of interest.

Approaches to the Evaluation of Intelligence Data

Workshop IST-036/RWS-005

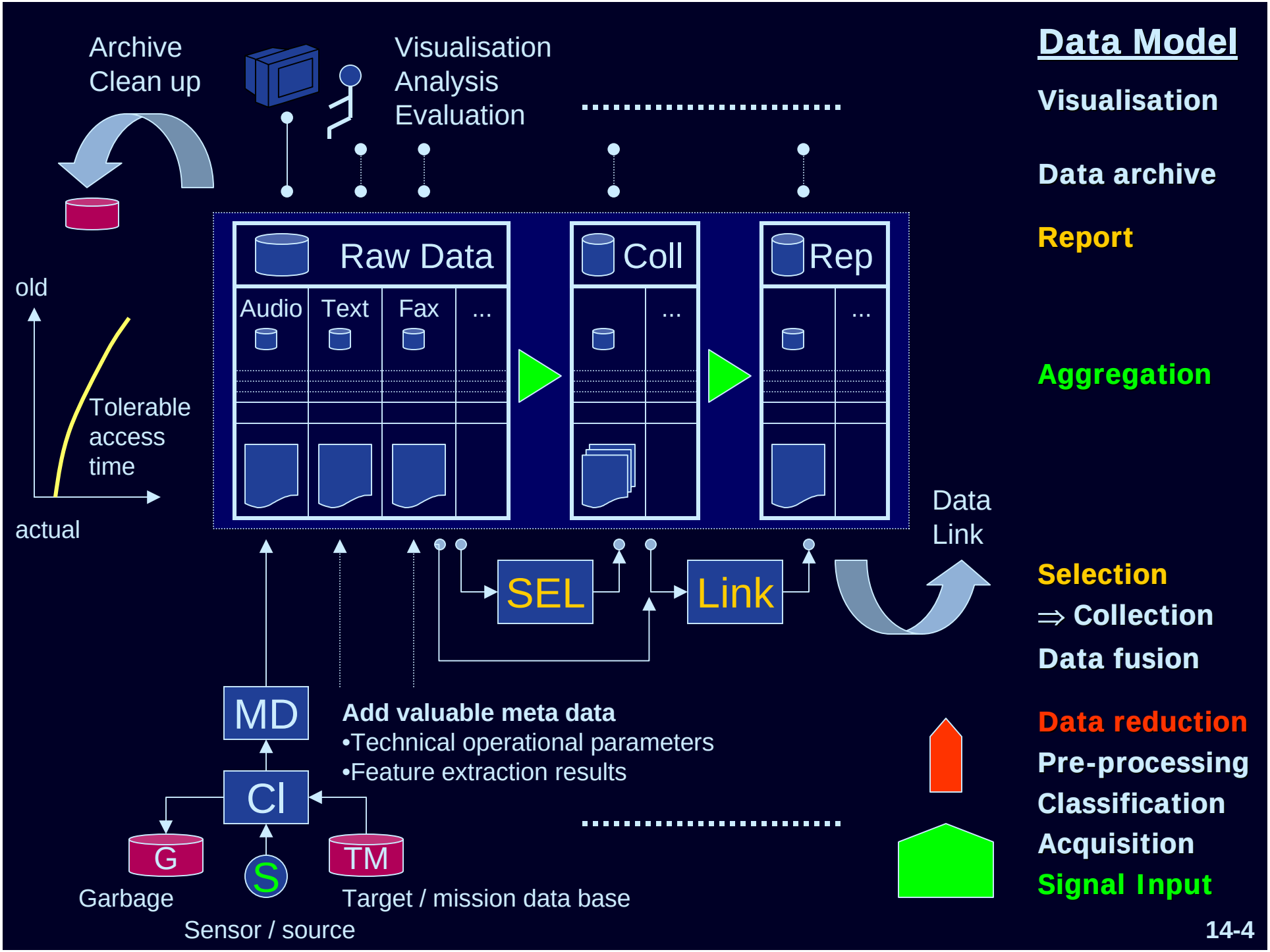
Massive Military Data Fusion and Visualisation: Users Talk with Developers

Definitions

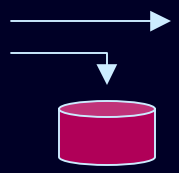
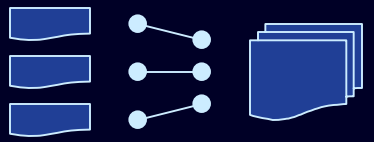
- Visualisation should generate an environment supporting the perception process of presented information.
- Visualisation should help average users to produce better results and help the ”artist” to become more creative.
- SIGINT makes contributions to describe the situation for military and political decision-makers on the basis of information intercepted from communication and signal distribution processes.

Overview

- Data model
- Information processing tasks
- Visualisation examples
- Further aspects


Archive
Clean up
Visualisation
Analysis
Evaluation
Raw Data
Audio
Text
Fax
...
Coll
...
Rep
...
old
Tolerable
access
time
actual
Data
Link
SEL
Link
MD
Add valuable meta data
•Technical operational parameters
•Feature extraction results
CI
G
S
TM
Garbage
Sensor / source
Target / mission data base
Data Model
Visualisation
Data archive
Report
Aggregation
Selection
⇒ Collection
Data fusion
Data reduction
Pre-processing
Classification
Acquisition
Signal Input

Information Processing Tasks

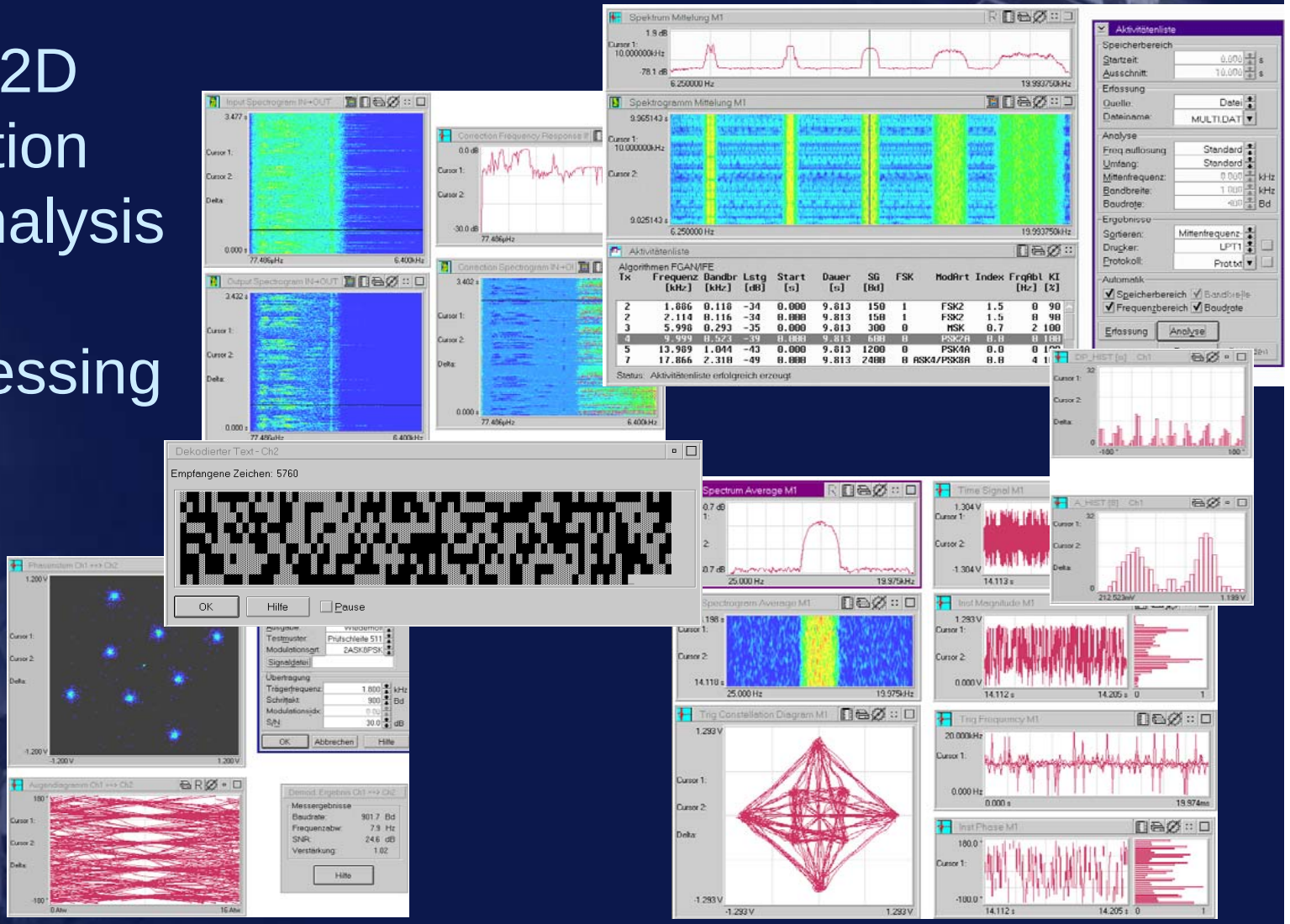
Task	Operation	Result	Example
Acquisition	Sensor Signal Processing in Realtime	Raw Signal File	■ Audio ■ Text ■ Fax
Classification	Determining targets of interest in respect of target/mission data base	 Signal of Interest Garbage	
Pre-processing	Adding technical operational parameters and additional features	 Additional Feature File	■ Time, duration ■ Frequency … ■ Language, speaker, topic, …
Selection ***	Interactive selection of a relevant signal collection		Signals within a window defined in parameter space
Reporting ***	Fillout a form or edit a result file !		

*** Visualisation needed !

Aggregation: Data Fusion – Reporting
Drill Down: ~ confidence, security, permission

Visualisation Examples: Raw Signal - Audio

Traditional 2D representation of signal analysis results incl. audio processing



Visualisation Examples: Raw Signal - Text

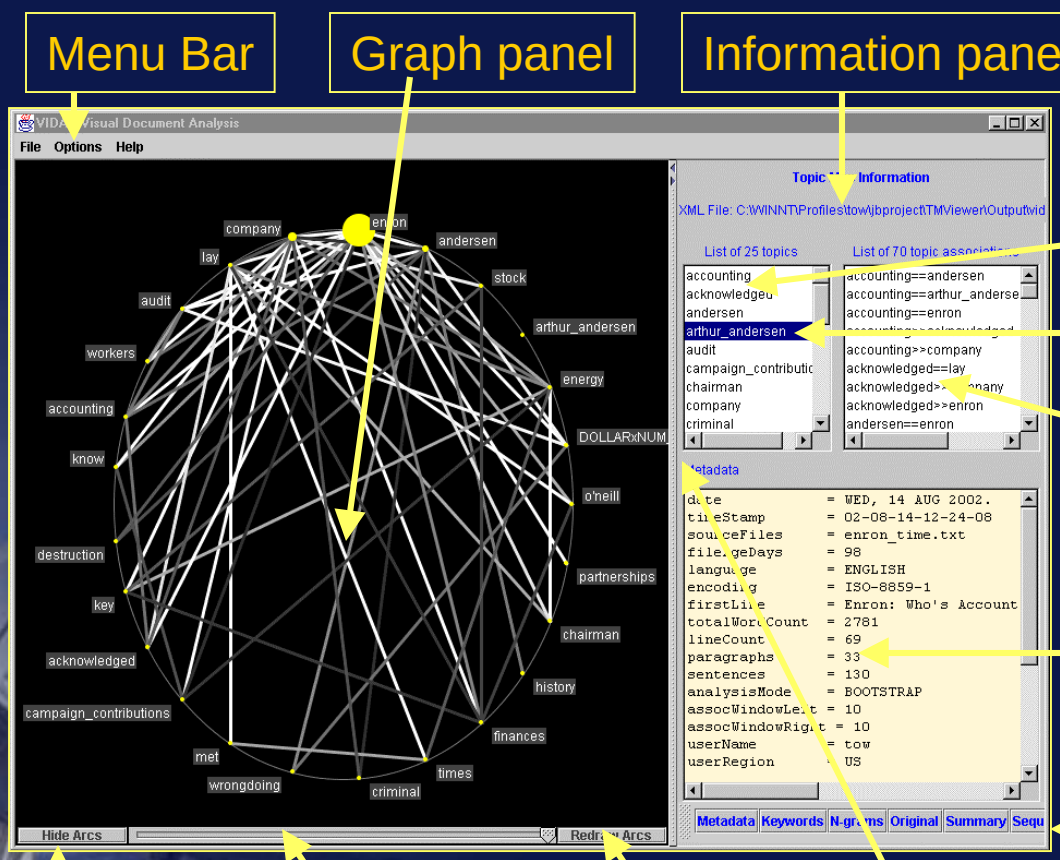
VIDA – Visual Document Analysis

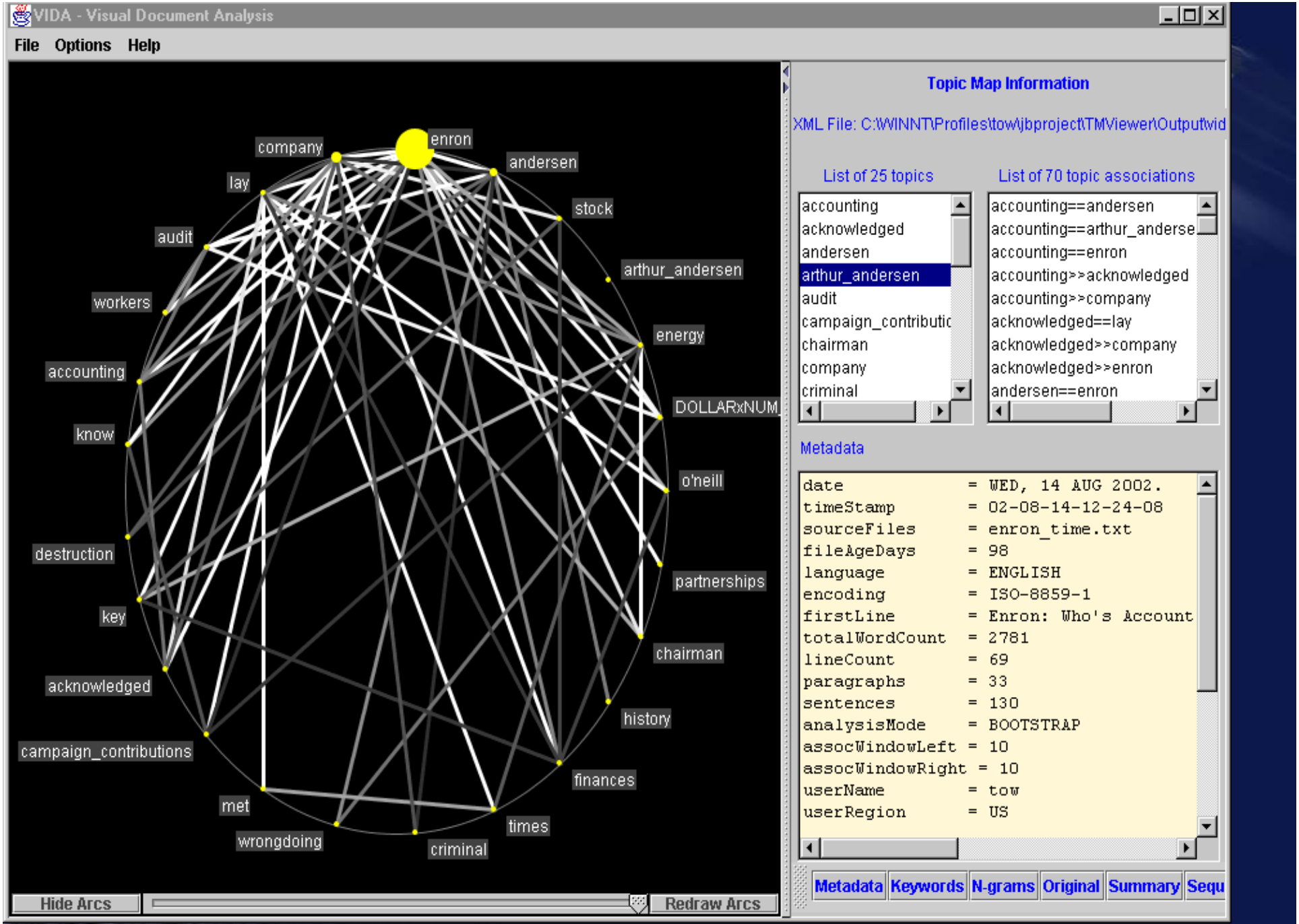
Dr. Michael Towsey, MEDAV

- Keyword Analysis
- Feature vector generation for automatic classification processes
- Grafical summary
- Textual summary

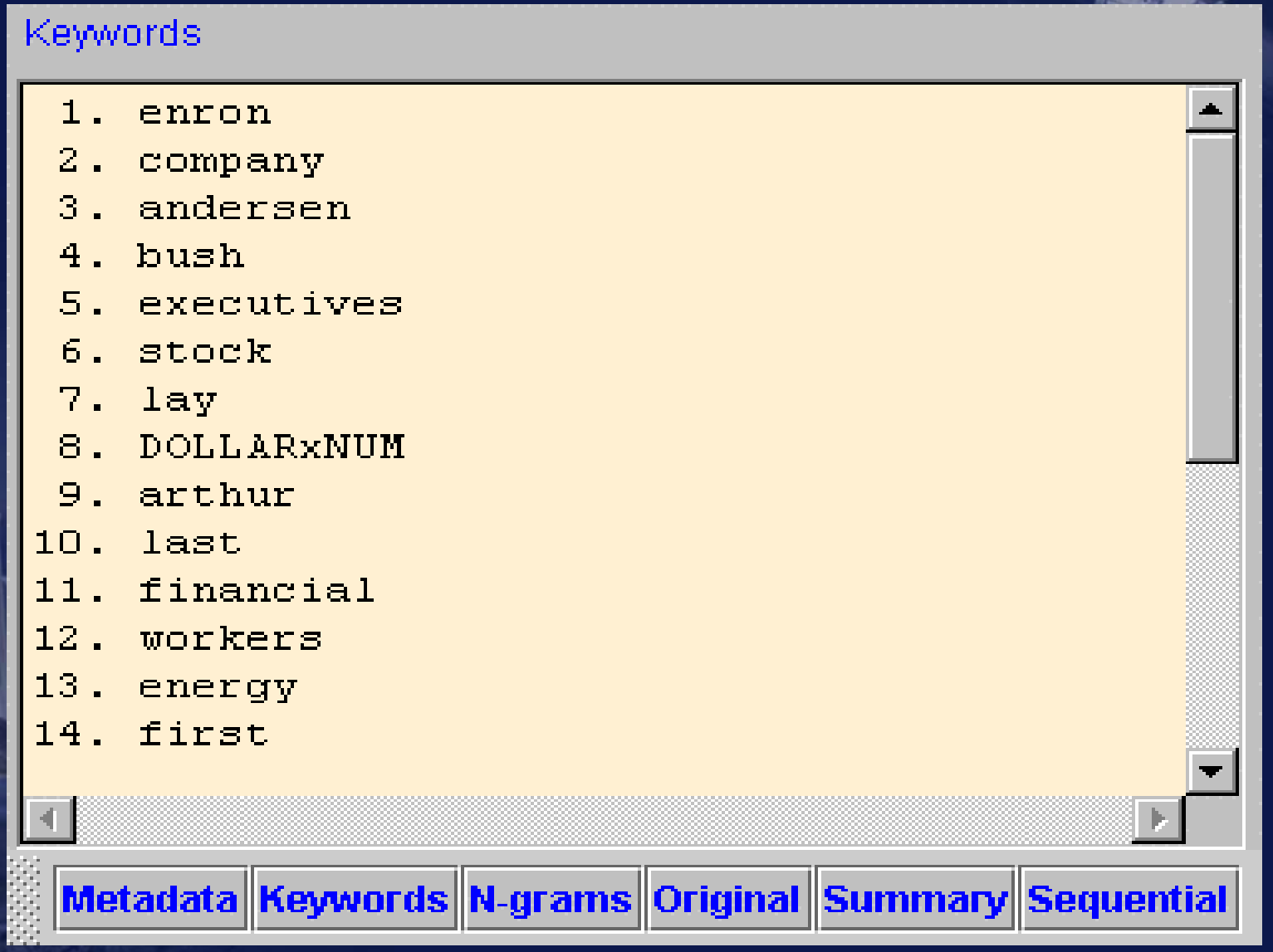
- Example: the Arthur Enderson / enron story

The full display




VIDA - Visual Document Analysis
File
Options
Help
enron
company
andersen
lay
stock
audit
arthur_andersen
workers
energy
accounting
DOLLARxNUM
know
o'neill
destruction
partnerships
key
chairman
acknowledged
history
campaign_contributions
finances
met
times
wrongdoing
criminal
Hide Arcs
Redraw Arcs
Topic Map Information
XML File: C:\WINNT\Profiles\tow\jbproject\TMViewer\Output\wid
List of 25 topics
accounting
acknowledged
andersen
arthur_andersen
audit
campaign_contributic
chairman
company
criminal
List of 70 topic associations
accounting==andersen
accounting==arthur_anderse
accounting==enron
accounting>>acknowledged
accounting>>company
acknowledged==lay
acknowledged>>company
acknowledged>>enron
andersen==enron
Metadata
date = WED, 14 AUG 2002.
timeStamp = 02-08-14-12-24-08
sourceFiles = enron_time.txt
fileAgeDays = 98
language = ENGLISH
encoding = ISO-8859-1
firstLine = Enron: Who's Account
totalWordCount = 2781
lineCount = 69
paragraphs = 33
sentences = 130
analysisMode = BOOTSTRAP
assocWindowLeft = 10
assocWindowRight = 10
userName = tow
userRegion = US
Metadata
Keywords
N-grams
Original
Summary
Sequ

Automatic Keyword Extraction



n-gram analysis

N-grams

TRIGRAMS (filtered)	FREQUENCY
oct 12 memo	3
attorney general john	2
BIGRAMS (filtered)	
arthur andersen	9
enron executives	7
DOLLARxNUM million	6
enron stock	4
last week	4
12 memo	3
bush administration	3
justice department	3
oct 12	3

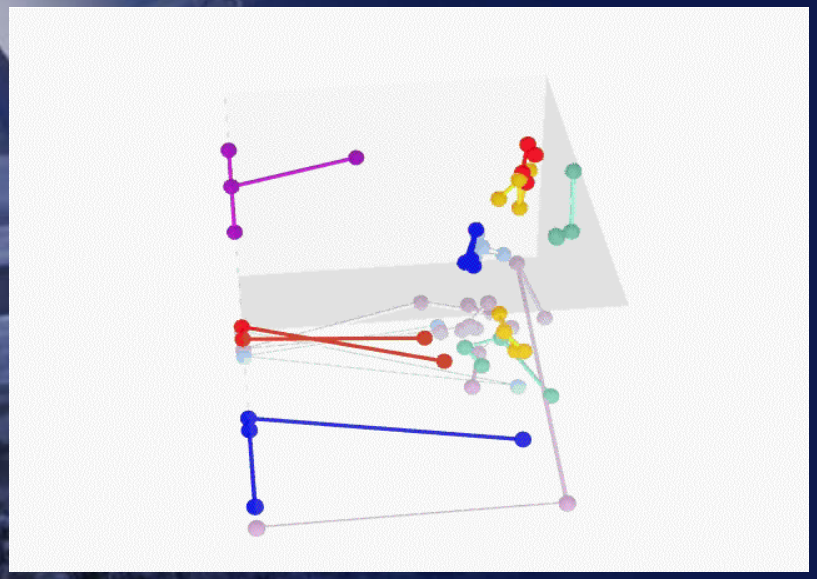
Metadata | Keywords | N-grams | Original | Summary | Sequential

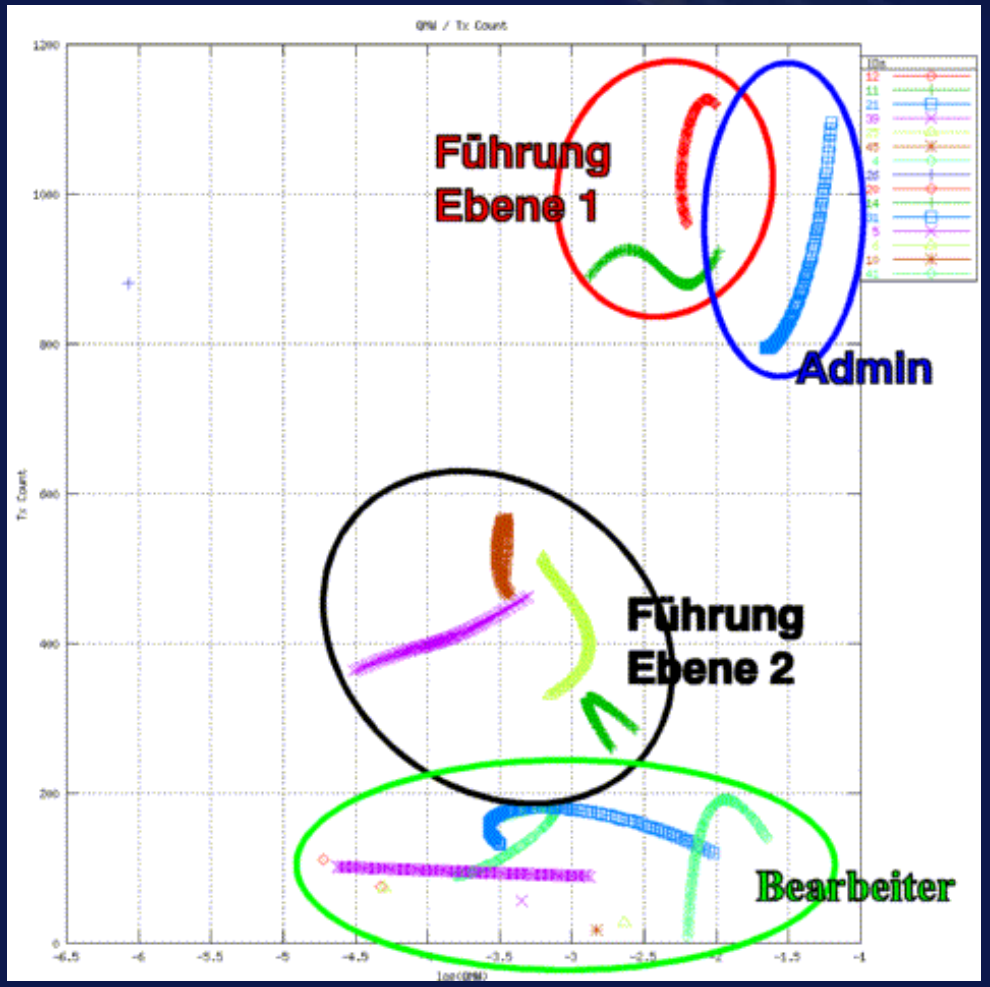
Association Analysis – "Arthur Anderson"

```
TOPIC= "arthur_andersen"

example          = woes workers who audited the company's book
example          = Supervisors at *Arthur* Andersen repeatedly
example          = Sources close to *Arthur* Andersen confirm
frequency        = 9
queryMatches     = arthur_andersen(9)
stopWords        = the(15) of(10) in(4) and(4) to(3) from(3) i

Ranked list of 30 non-stop words associated with topic word:-
     Word        assocFreq offset  SD  weight relWeight
  1. enron               4  +2.0   5.2  2.07  1.00
  2. destruction         3  -3.3   9.8  1.02  0.49
  3. criminal            1  +1.0   0.0  1.00  0.48
  4. confirm             1  +1.0   0.0  1.00  0.48
  5. bankrupt            1  -1.0   0.0  1.00  0.48
```

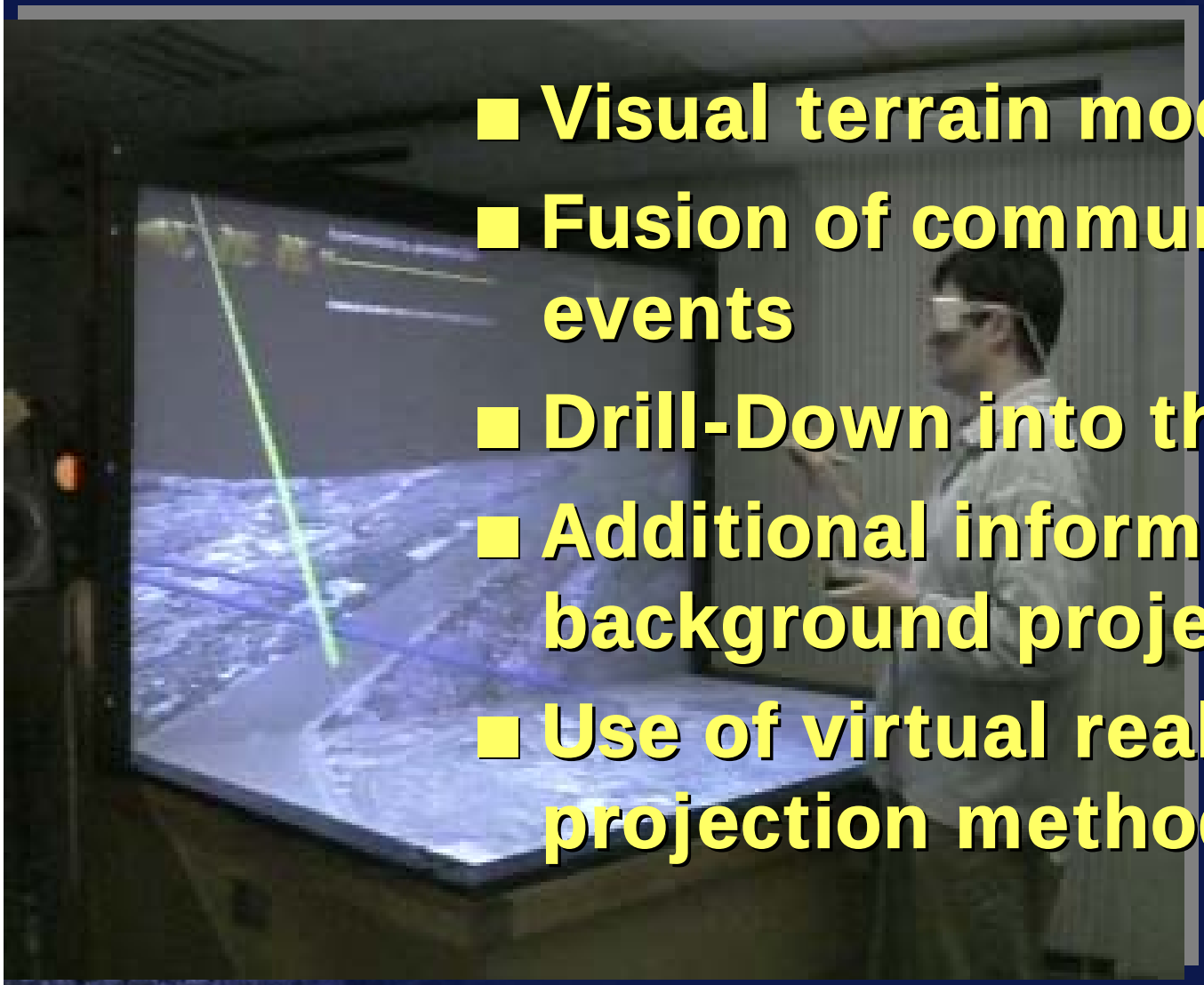
Visualisation Examples: Collection – E-mail

- Command & Control structures
- Using statistical methods of pattern recognition



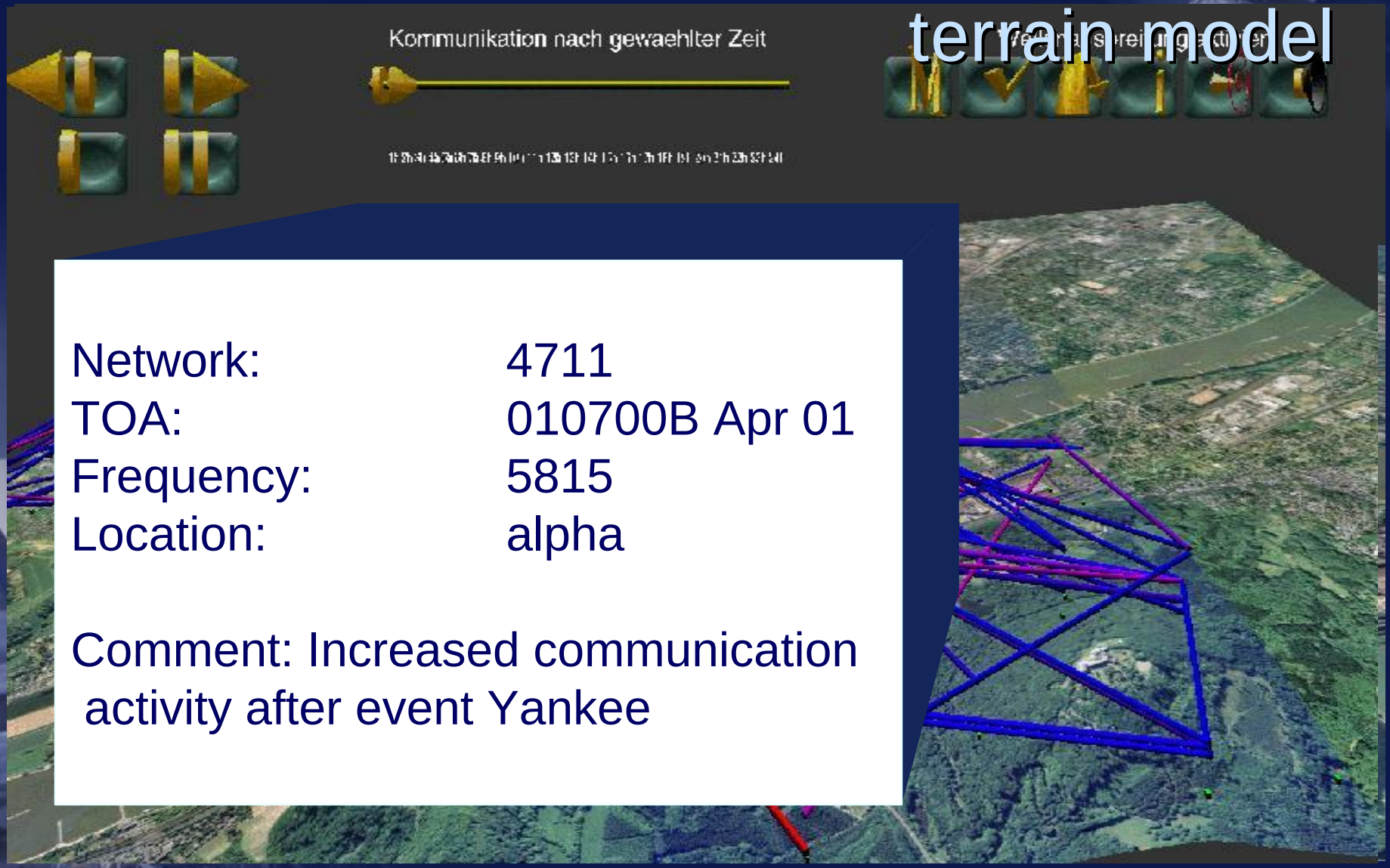


Visualisation Examples: Collection – Radio network I



- Visual terrain models
- Fusion of communication events
- Drill-Down into the raw data
- Additional information by background projection
- Use of virtual reality projection methods

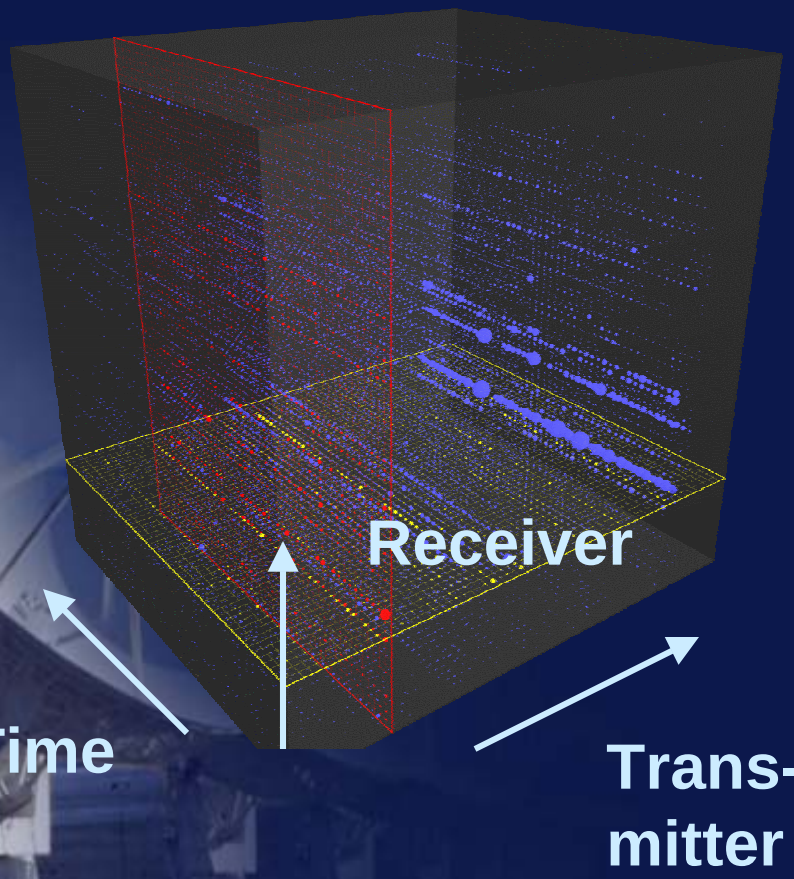
COMINT-Situation displayed in a terrain model



Network: 4711
TOA: 010700B Apr 01
Frequency: 5815
Location: alpha

Comment: Increased communication activity after event Yankee

Visualisation Examples: Collection - Radio network II



- **technical parameters:**
  - Call sign sender
  - Call sign receiver
  - Time of Arrival

Frequency of contact is described by the size of the sphere